Technical Bulletin

Part No. 74-0130

DataStage Sybase IQ12 Load

This technical bulletin describes Release 1.2 of the DataStage Sybase IQ12 Load stage. This stage loads data from DataStage into the Sybase IQ indexing engine for Version 12.

Printing History

First Edition (74-0130) for Release 1.0, March 1999
Second Edition (74-0130) for Release 1.1, March 1999
Third Edition (74-0130) for Release 1.2, July 2000
Updated for Release 1.2, August 2002
Updated for Release 1.2, August 2003

How to Order Technical Documents

To order copies of documents, contact your local Ascential subsidiary or distributor, or call our main office at (508) 366-3888.

Documentation Team: Marie E. Hedin

Introduction

This technical bulletin describes the following for Release 1.2 of the DataStage Sybase IQ12 Load stage, updated for DataStage Release 7.0:

- Functionality
- Terminology
- Installation
- Index sets
- Disk overflow handling
- Stage and link properties

The Sybase Adaptive Server IQ for Version 12 (Sybase IQ12) is an advanced indexing engine, not a database management system. The goal of the Sybase IQ12 Load stage is to enable DataStage Release 7.0 to rapidly and efficiently load data into existing Sybase IQ index sets.

Each input link to the stage represents a stream of rows to load into a Version 12 Sybase Adaptive Server IQ index set or joined index set. Reference links and output links have no meaning in the context of the Sybase IQ12 stage and are not allowed.

The Sybase IQ12 Load stage does not replace the Sybase IQ Loader plug-in used with versions of Sybase IQ before Version 12. Continue to use Sybase IQ Load to do so. You can create jobs with both the Sybase IQ12 Load and the Sybase IQ Load stages if you need to load different versions of Sybase IQ within the same job. The Sybase IQ12 Load stage cannot load Sybase Adaptive Server IQ versions before Version 12. Use the Sybase IQ Load stage to do this.

TheSybase IQ12 Load is not a migration tool. Use Sybase tools to generate scripts if you need to migrate data from earlier versions of the Sybase Adaptive Server IQ. Consult the Sybase Adaptive Server IQ Version 12 documentation for help with migration issues.

Functionality

The Sybase IQ12 Load stage has the following functionality:

- Generation and optional automatic execution of the Sybase IQ Version 12 commands to load indexes with data from input links.
- New Sybase IQ12 commands follow the SQL standard to use tables instead of indexes. Old Sybase IQ commands are no longer supported.

- Two methods of loading an index set: manual loading or automatic loading using Open Client/Open Server (OCOS) or Open Database Connectivity (ODBC).
- Simplified loading of joined index sets that does not require a specific sequence for loading data. You can use the SYNCHRONIZE command to build joined indexes rather than a batch file.
- Support for the server as a stand-alone database without needing a catalog server or direct use of an SQL server. This results in sophisticated indexing and query optimization.
- The ability to load tables during execution of queries.
- Automatic generation of overflow data files if the first data file exhausts physical disk space.
- Support for data files that exceed the 2-GB file size limit for 64-bit file systems.
- The ability to specify Sybase IQ12 commands to be run before and after the insert operation to send diagnostic or verification output to the DataStage log.
- Generation of data files in delimiter-separated ASCII format.
- Support for NLS (National Language Support). For more information, see *DataStage NLS Guide*.

The following functionality is not supported:

- Deletion and recreation of the index set itself.
- Other modes of operation supported in database load utilities, for example, update existing rows only.
- Automatic execution of load commands when the Sybase IQ server resides on a machine other than the one for the DataStage job.
- Loading of joined index sets can only be done using an after-job subroutine, because the stage instance cannot guarantee that the columns of different tables in the index set are loaded in the correct order. (The job compiler may draw a process boundary through a given stage instance so that there is no one process address space that knows the status of all the links connected to the stage.)
- Use of named pipes to load data.

- Support for loading of Sybase IQ before Version 12. Use Sybase IQ Load to do this.
- Meta data import.

Added Functionality

The following functionality is changed or added from that in the Sybase IQ Load stage. The prompt is included in parentheses whenever the program-visible property name is referenced. See "Properties" on page 9 for more information.

- The DELETE statement replaces the IQ DELETE command. The IQDELWITH property (IQ DELETE…WITH) no longer exists since the DELETE statement does not support a WITH clause. The DODELETE property (Clear Before Load) contains a new list box choice of Yes/No. The user-visible prompts and property names for the IQDELFROM (IQ DELETE…FROM) and IQDELWHERE (IQ DELETE…WHERE) properties no longer use the IQ prefix.
- The LOAD TABLE statement replaces the IQ INSERT command. The LOAD TABLE statement has about 20 optional clauses. To support these optional clauses, the IQLOADWITH property (IQ DELETE…WITH) is changed to LOADOPTIONS (LOAD TABLE…OPTIONS). This lets you specify any of the optional clauses for the LOAD TABLE statement in addition to the WITH clause. Also added are the FORMAT, STRIP, and CHECKPOINT properties to control various LOAD TABLE options.
- The PASSWORD property is encrypted.
- The AUTOLOAD property (Load Automatically), renamed LOADMETHOD, contains a new list box with choices of Manual, Autoload via OCOS, and Autoload via ODBC. These changes also allow the later addition of the named pipe loading methods.
- The run-time command processor *dbisql* replaces *isql* in all the documentation and script files.
- The grid property term *table* replaces the term *indexes*.
- Loading a joined index set is simplified to change how Sybase IQ12 Load loads these structures. Since the data for each joined index set no longer needs to be loaded in a specific sequence, the joined index set no longer needs to be loaded outside the job itself. You can load joined indexes using Automatic loading. However, you must now explicitly execute a SYNCHRONIZE command after the job completes (for example, as an after-job command). Also, the notion of a load sequence number in the

property list and the logic associated with it in the creation and execution (IQLOAD.BAT) of a joined index set load are removed. The properties IQJOINIDXSET and IQLSN are removed.

- Since synchronizing join indexes can be a time-consuming process, the synchronization runs after the job completes (as opposed to after each stage completes a series of inserts and deletes to a joined index). A batch file is no longer needed. You can run the synchronization as an after-job subroutine.
- The property CTRLFILE (Control File Name) is renamed SQLFILE (SQL File Name) and the stage generates *.sql* files to contain the IQ12 load commands.

Terminology

The next two sections explain Sybase IQ12 and the Sybase IQ12 Load stage terms used in this document.

Sybase IQ12 Terminology

The following table lists the Sybase IQ12 terms used in this document:

Term	Description
DELETE	The Sybase IQ12 command used to delete rows from a table.
index	A single column in an index set.
index set	The Sybase IQ12 equivalent of a table. It is a collection of named, typed indexes on columns of data which may have come from a Sybase SQL Server database, a foreign database, or a flat file. The Sybase IQ12 Load stage loads data into index sets. Every index set definition has associated with it a Sybase SQL Server table definition, because Sybase IQ12 uses SQL Server to catalog information about its index sets.
index space	The Sybase IQ12 entity that contains index sets. The index space owns disk and other resources, and provides a handle for administration.
IQ Server	The server engine to which you connect in order to use Sybase IQ12. An IQ Server instance provides access to one or more index spaces.
joined index set	A set of indexes created to allow a relational join between two or more tables.

Term	Description
LOAD TABLE	The Sybase IQ12 command used to load data into a table.
SYNCHRONIZE	The Sybase IQ12 command used to synchronize joined indexes.

Sybase IQ12 Load Terminology

The following table lists the Sybase IQ12 Load terms used in this document:

Term	Description
Load stage	A passive stage whose role in a DataStage job is to take streams of tabular data and load them into tables of a target database.
data file	An ASCII file of row and column data from an input link that is to be loaded. The Sybase IQ12 Load stage generates these files with noncharacter columns separated by vertical bars (|). "CHAR delimiter" specifies the character to use to separate character columns. If "CHAR delimiter" does not have a value, character columns are written in fixed-width format.
SQL file	A control file of Sybase IQ12 commands that loads or reloads an index set. A DataStage job generates one SQL file for each input link to each instance of the Sybase IQ12 Load stage. Control files can be executed by piping them to the Sybase *dbisql* utility.
stage instance	An individual stage of a given type, appearing as an icon in a job design.

Installing the Plug-In

For instructions and information supporting the installation, see *DataStage Plug-In Installation and Configuration Guide*.

Index Sets

The next two sections describe loading index sets and joined index sets.

Loading Index Sets

Sybase IQ12 Load supports the following methods using the Load Method property to load the data from its input links into Sybase IQ12 index sets:

- Manual loading
- Automatic loading

Manual Loading

The Sybase IQ12 Load stage instance generates an ASCII data file and an SQL control file for each input link, but does not load the data into Sybase IQ12. You can load the data later by redirecting the control file to the Sybase *dbisql* utility. Manual loading is the default.

Automatic Loading

The methods to load data automatically are:

- OCOS
- ODBC

Automatic loading works like manual loading. Rows arriving at an input link are written to a data file as in manual loading. When the link reaches the end of the data, appropriate DELETE and LOAD TABLE commands are generated and executed using a Sybase Client-Library connection to the IQ Server 12. The commands to be executed are also written to an SQL file to log the activity.

OCOS. Automatic loading via OCOS works when both the DataStage server and the IQ Server12 reside on the same machine. It also works if the following is true:

- The two servers reside on different machines that are connected by a local area network (LAN).
- Open Client is installed on the machine hosting the DataStage server.
- Both machines share a common directory for the output file.

ODBC. Automatic loading via ODBC works if both the DataStage server and IQ Server 12 reside on the same machine. It also works if the following is true:

- The two servers reside on different machines over a LAN.
- The IQ Server 12 client is installed on the machine hosting the DataStage server.
- Both machines share a common directory for the output files.

For more information about overflow directories, see “Disk Overflow Handling” on page 7.

Loading Joined Index Sets

Joined index set loading has been simplified. Since the data for each joined index set no longer needs to be loaded in a specific sequence, you can load joined indexes automatically. You need to explicitly execute a SYNCHRONIZE command after job completion. You can use the Post-insert Command property or ExecDOS to do this.

The data for each table in the join must be synchronized to allow you to bring the joined indexes up to date making them available for queries to use.

Support for joined indexes in the Sybase IQ12 Load stage is complicated because the input links to a stage instance are not guaranteed to run in the same process. Depending on the overall design of the job, the DataStage job compiler may draw process boundaries through an Sybase IQ12 Load stage instance. This makes it impossible to know while the job is running when the last link has closed. Consequently, the actual synchronizing of data in a joined indexes has to be done outside the job itself. You can execute the SYNCHRONIZE command as an ExecDOS after-job subroutine to load and synchronize joined indexes.

To load and synchronize joined index sets:

1. Open the job in the DataStage Designer.
2. Choose **Edit ➤ Job Properties** to display the **Job Properties** dialog box. From here you can configure job parameters.
3. On the **General** page, select **ExecDOS** from the **After-job subroutine** list.
4. Enter the following command in the **Input Value** field to run SYNCHRONIZE against the appropriate tables:

```
dbisql -q -c Userid=dba;
Password=sql;ServerName=asiqdemo;DatabaseName=asiqdemo;
SYNCHRONIZE JOIN INDEX emp_dept_join1, emp_dept_join2
```

Although the synchronizing must be done outside the job itself, the actual loading of a joined index can be done using any loading method. Since the data for each joined index no longer needs to be loaded in a specific sequence, it is possible to load joined indexes using automatic loading. You must explicitly execute a SYNCHRONIZE command after the job completes.

Disk Overflow Handling

The Sybase IQ12 Load stage must be able to handle load operations in the multiple-gigabyte range. These large data sets can exhaust the free space on the disk drive or partition that is receiving the data file.

You can use a semicolon-separated list of directory paths as the value of the Output Path property to handle these situations. If the stage runs out of disk space during a job run and cannot write a row to the data file, it opens an overflow data file in the second directory in the list and continues. In this way, the data can be spread among multiple disk drives or partitions.

Add the pathnames of overflow files to the FROM clause of the LOAD TABLE command that loads the data.

Properties

The Sybase IQ12 Load stage supports stage and link properties that are visible from the DataStage Designer. You need to supply values for these properties in the stage grid-style editor. For more information about the grids, see Appendix A in *DataStage Designer's Guide*.

The tables in the next two sections include the following column heads:

- **Prompt** is the text that you see in the stage editor user interface.
- **Type** is the data type of the property.
- **Default** is the value used if you do not supply a value.
- **Description** gives details about the properties.

Stage Properties

The Sybase IQ12 Load stage supports the following stage properties. The stage properties are listed in the logical order in which you design a job.

Stage Properties

Prompt	Type	Default	Description
Load Method	List	Manual	The method in which the data is to be loaded into the Sybase IQ12 tables. Valid methods are: **Manual.** Generates the SQL and data files which must be manually loaded using *dbisql*. **Autoload via OCOS.** Automatically loads the data using Open Client/Open Server. **Autoload via ODBC.** Automatically loads the data using ODBC. Automatic loading tells the Sybase IQ12 Load stage to connect to the IQ server and execute the commands in the SQL file after the last row of data has been written to the corresponding data file. If this option is used for joined indexes, you must explicitly execute SYNCHRONIZE commands to update the joined index. (Manual/Autoload via OCOS/Autoload via ODBC)

Stage Properties (Continued)

Prompt	Type	Default	Description
IQ Server/ Datasource Name	String	None	The name of the IQ12 Server or ODBC data source as defined for any IQ client program. On Windows NT platforms, Autoload using OCOS uses the name of the IQ12 Server defined in *sql.ini.* On UNIX platforms, Autoload using OCOS uses the name of the IQ12 Server defined in *$sybase/interfaces*. Autoload using ODBC uses the name of the ODBC database. Required for autoloads.
IQ Database Name	String	None	The name of the target Sybase IQ12 database. This name appears as the argument to a connection request. Required for autoloads.
IQ User ID	String	None	The Sybase IQ12 user name used when connecting to the IQ12 Server to perform the load. Required for autoloads.
IQ Password	String	None	The Sybase IQ12 password used when connecting to theIQ12 Server to perform the load. Required for autoloads.
Output Path	String	C:\temp	Semicolon-separated lists of absolute pathnames where the Sybase IQ12 Load stage creates SQL and data files. SQL files are lists of IQ12 commands always written to the first directory in the path. Data files are ASCII files of data from input links created first in the first directory in the path. If the data overflows the disk space in the current directory, an overflow data file is created in the next directory in the list, and so on until the end of the data or the disk space is exhausted.

Link Properties

The Sybase IQ12 Load stage supports the following input link properties, which you use to write to Sybase tables. The properties are listed in the logical order in which you design a job.

Link Properties

Prompt	Type	Default	Description
Table Name	String	None	The name of the target Sybase IQ12 table to be loaded with data from this link. This name is the argument of the LOAD TABLE command. Required.

Link Properties (Continued)

Prompt	Type	Default	Description
SQL File Name	String	Table-name.sql	The SQL file (generated by the Sybase IQ12 Load stage) containing deletion and insertion commands to run at database load time. This file is always created for documentation purposes, even if the IQ12 commands are executed using OCOS or ODBC. Its name defaults to the table name, with an extension of *.sql*. The SQL file is always created in the first directory in "Output Path."
Data File Name	String	Table-name.dat	The file name of a flat ASCII output file generated by Sybase IQ12 Load containing the rows and columns of data to load into the table for the link. The file name defaults to the table name, with an extension of *.dat*. The data file is created in the first output path directory. If disk space under this directory is exhausted, data overflows into a new data file of the same name in the second output path directory, and so on until the end of the data is reached or the last output path directory is exhausted.
CHAR Delimiter	String	| (vertical bar)	The character string that delimits CHAR and VARCHAR columns in the data file. Noncharacter columns are always delimited by vertical bars (|), regardless of the CHAR delimiter.
Clear Before Load	List	Yes	Controls whether a DELETE command is generated before the LOAD command that loads the new data into the table. For single tables, an DELETE command is generated. In both cases, the FROM and WHERE clauses in the corresponding properties are added to the command. (Yes/No)
DELETE...FROM	String	None	This permits full specification of the FROM table-list clause of the DELETE command. Defaults to the table name (the value of the TABLE property for the link). The keyword FROM is not required. The absence of a FROM clause is equivalent to the TRUNCATE command.
DELETE... WHERE	String	None	Specifies optional *search_conditions* for the DELETE command. If defined, the *search_conditions* go into the WHERE clause of the command. The keyword WHERE is not required.

Link Properties (Continued)

Prompt	Type	Default	Description
LOAD TABLE... FORMAT	List	ascii	Specifies the optional format option for the LOAD TABLE command. (ascii/binary)
LOAD TABLE... STRIP	List	ON	Specifies the optional strip option for the LOAD TABLE command. (ON/OFF)
LOAD TABLE... CHECKPOINT	List	OFF	Specifies the optional checkpoint option for the LOAD TABLE command. (ON/OFF)
LOAD TABLE... LOADOPTIONS	String	None	Specifies load options for the LOAD TABLE command.
Pre-insert Command	String	None	The full literal text of an optional IQ12 command to run before the DELETE and LOAD TABLE commands that make up the load. This can be used to run IQ12 diagnostic commands. Output from these commands appears in the DataStage job log during automatic loading.
Post-insert Command	String	None	The full literal text of an optional IQ12 command to run after the DELETE and LOAD TABLE commands that make up the load. This can be used to run IQ12 diagnostic commands. Output from these commands appears in the DataStage job log during automatic loading.